PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-170552
(43)Date of publication of application : 17.06.2004

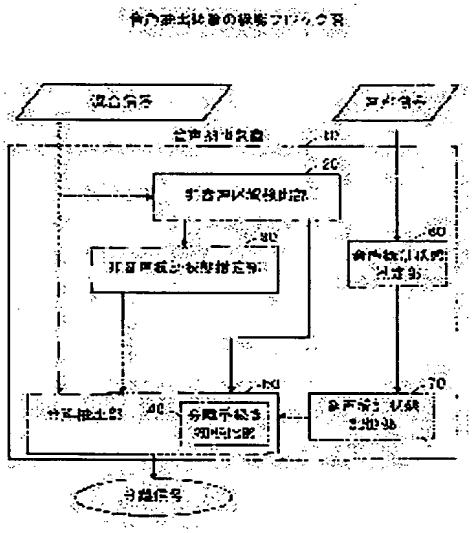
(51)Int.Cl. G10L 15/20
G10L 11/02
G10L 15/00
G10L 15/02
G10L 15/04
G10L 21/02

(21)Application number : 2002-334276
(22)Date of filing : 18.11.2002

(71)Applicant : FUJITSU LTD
(72)Inventor : ENDOU KAORI
OTA TAKASHI
SASAKI HITOSHI
MATSUBARA MITSUYOSHI

(54) SPEECH EXTRACTING APPARATUS

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a speech extracting apparatus that extracts a speech of a specified person from a mixed signal of a speech signal of the specified person and a non-speech signal other than the speech signal while taking a statistical state of a sound source depending upon time into consideration by using a plurality of input signals in which a plurality of sound sources are mixed and the statistical independence of the sound sources included in the input signals.
SOLUTION: A non-speech section detection part 20 detects a non-speech section including no speech from a mixed signal in which a plurality of sound sources are mixed, a non-speech statistical state estimation part 30 estimates a non-speech statistical state by using information on the detected non-speech section, and a speech extraction part 50 extracts at least one speech by sequentially updating the mixed signal according to an initial procedure selected by a separation procedure initialization part 40 by using the statistical independence of the estimated non-speech statistical state and a statistical state, regarding the speech, stored in a speech statistical state storage part 70.



LEGAL STATUS

[Date of request for examination] 12.05.2005
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2004-170552
(P2004-170552A)
(43)公開日　平成16年6月17日(2004.6.17)

(51)Int.Cl.[7]	F I			テーマコード（参考）
G10L 15/20	G10L	3/02	301C	5D015
G10L 11/02	G10L	3/00	513C	
G10L 15/00	G10L	9/00	301A	
G10L 15/02	G10L	3/00	551A	
G10L 15/04	G10L	3/02	301Z	

審査請求 未請求　請求項の数 5　OL　（全 14 頁）　最終頁に続く

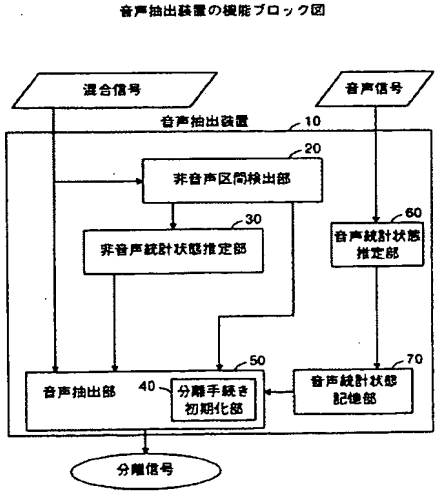
(21)出願番号　特願2002-334276(P2002-334276)
(22)出願日　平成14年11月18日(2002.11.18)

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　遠藤　香緒里
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　大田　恭士
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　佐々木　均
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

最終頁に続く

(54)【発明の名称】音声抽出装置

(57)【要約】
【課題】特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出すること。
【解決手段】非音声区間検出部２０は、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、非音声統計状態推定部３０は、検出された非音声区間の情報を用いて非音声の統計状態を推定し、推定された非音声の統計状態と音声統計状態記憶部７０において知識として記憶されている音声に関する統計状態との統計的な独立性を用いて、音声抽出部５０は、分離手続き初期化部４０よって選択された初期手続きから、混合信号を逐次更新することによって少なくとも一つの音声を抽出する
【選択図】　図２



【特許請求の範囲】
【請求項１】
特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、
複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段と、
前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段と、
非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段と、
を備えたことを特徴とする音声抽出装置。
【請求項２】
前記非音声区間検出手段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする請求項１に記載の音声抽出装置。
【請求項３】
前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする請求項１に記載の音声抽出装置。
【請求項４】
前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする請求項１に記載の音声抽出装置。
【請求項５】
前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段を備えたことを特徴とする請求項１に記載の音声抽出装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、複数音源が混合した複数の入力信号と、該入力信号に含まれる音源の統計的な独立性を用いて、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出する音声抽出装置に関し、特に、携帯電話システムおよび実況中継において、混入した環境雑音と送話音声を分離し、送話音声のみを送信することができる音声抽出装置に関する。
【０００２】
【従来の技術】
複数の信号源から出た信号が混合している時、異なる信号源からでた信号が互いに統計的に独立であるという条件のみから、各々の信号を分離する方法は独立成分分析と呼ばれている。現在、よく知られている独立成分分析の手法（例えば、非特許文献１参照）では、信号源が未知で信号源の信号が一定の割合で線形的に結合され、観測されると仮定し、統計的な独立性の概念を基に混合信号を分離する手法を用いる技術が開示されている。
【０００３】
また、独立成分分析において、混合信号の数が統計的に独立な信号源の個数と等しいか、あるいはそれ以上の場合、該独立信号の混合である混合信号を分離し、未知である信号源の信号を特定する方法は知られている。例えば、非特許文献１がある。しかしながら、非特許文献による手法では混合信号の数より多い統計的に独立な音源信号からなる混合信号を分離することはできない。

【０００４】
これに対して、特許文献１は混合信号の数より多い独立な音源信号を混合信号の観測時間を遅延し、混合信号を逐次的に分離更新する方法を開示している。
【０００５】
【非特許文献１】
Ｓ．Ａｍａｒｉ，、Ｔ．Ｐ．Ｃｈｅｎ、およびＡ．Ｃｉｃｈｏｃｋｉ共著論文「Ｓｔａｂｉｌｉｔｙ ａｎａｌｙｓｉｓ ｏｆ ｌｅａｒ ｎｉｎｇ ａｌｇｏｒｉｔｈｍｓ ｆｏｒ ｂｌｉｎｄ ｓｏｕｒｃｅ ｓｅｐａｒａｔｉｏｎ」Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋｓ，Ｖｏｌ．１０，Ｎｏ．８，ｐｐ．１３４５－１３５１
【特許文献１】
特開２００２－５５９６９号公報
【０００６】
【発明が解決しようとする課題】
しかしながら、この特許文献１の技術では、混合信号を逐次的に分離更新する方法に時間に依存した統計的機構が含まれていないため、音源が移動している状況などを考えれば、信号源の混合のされ方自体が時間変化する状況では混合信号を適切に分離することが望めない。
【０００７】
このため、混合信号の数以上の統計的に独立な音源信号からなる混合信号を、時間によって変動する混合信号の統計状態を考慮して、混合信号を逐次的に分離しつつ、目的に応じた音源信号を抽出することは、極めて重要な課題となっている。特に、携帯電話システムおよび実況中継においては、話者がマイクを持って移動する可能性があるため、環境雑音が話者の所有するマイクに混入する可能性がある。このような状況において、時間変化する環境雑音を考慮して、混入した環境雑音と送話音声を分離し、送話音声のみを送信する必要性がある。
【０００８】
この発明は、上記従来技術による課題を解決するためになされたものであり、複数音源が混合した複数の入力信号と、該入力信号に含まれる音源の統計的な独立性を用いて、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出する音声抽出装置を提供することを目的とする。
【０００９】
【課題を解決するための手段】
本発明は、上記目的を達成するためになされたものであり、請求項１に係る音声抽出装置は、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段（図２に示す非音声区間検出部２０に対応する）と、前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段（図２に示す非音声統計状態推定部３０に対応する）と、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段（図２の音声抽出部５０に対応する）と、を備えたことを特徴とする。
【００１０】
この請求項１の発明によれば、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、検出された非音声区間の情報を用いて非音声の統計状態を推定し、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の声を抽出することとしたので、時間変化する環境雑音を伴う混合信号から目的とする音声信号を抽出することができる。
【００１１】
また、請求項２に係る音声抽出装置は、請求項１の発明において、前記非音声区間検出手

段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする。

【００１２】

この請求項２の発明によれば、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することとしたので、単位時間毎に変化する環境雑音を非音声信号として抽出でき、該抽出結果を用いて、非音声信号の統計状態を推定することができる。

【００１３】

また、請求項３に係る音声抽出装置は、請求項１の発明において、前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする。

【００１４】

この請求項３の発明によれば、単位時間毎に混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうこととしたので、単位時間毎に混合信号の中から環境雑音を抽出し、単位時間毎に環境雑音の統計状態を推定することができる。

【００１５】

また、請求項４に係る音声抽出装置は、請求項１の発明において、前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする。

【００１６】

この請求項４の発明によれば、推定された時間によって変動する非音声の統計状態と知識としての音声に関する統計状態との統計的な独立性を用いて、非音声の統計状態と音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、単位時間毎に逐次更新する分離手続きを行うこととしたので、混合信号の統計状態から音声信号の統計状態を分離でき、よって混合信号から目的とする音声信号を抽出することができる。

【００１７】

また、請求項５に係る音声抽出装置は、請求項１の発明において、前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段（図２の分離手続き初期化部４０に対応する）を備えたことを特徴とする。

【００１８】

この請求項５の発明によれば、混合信号の分離手続きを初期化することとしたので、手続きの初期状態から安定であり確度の高い音声信号の抽出ができる。

【００１９】

【発明の実施の形態】

以下に添付図面を参照して、この発明に係る音声抽出装置の好適な実施の形態を詳細に説明する。なお、本実施の形態では、便宜上、２つの混合信号を分離する場合について説明するが、少なくとも一つの音声信号源並びに少なくとも一つの非音声信号源からなる混合信号を分離する場合についても、目的とする音声信号の抽出が可能である。

【００２０】

以下、本発明の概略構成を説明した後、音声抽出装置の処理手順の概念を説明する。最後に、後音声抽出装置の処理手順としていくつかの実施の形態をフローチャートに基づいて説明する。

【００２１】

まず、本発明である音声抽出装置の概略構成を説明する。図１は本発明の概略構成を示す図である。複数の音源として、非音声音源１から非音声音源ｎまでのｎ個の音源と音声音源１から音声音源ｍまでのｍ個の音源から、各音声音源に対応する音声を抽出するための

音声抽出装置１から音声抽出装置ｍまでがある。各音声抽出装置にはマイクが備わっておりここから非音声音源および音声音源から発信される混合信号を受信する。また、各音声抽出装置は、受信した混合信号から目的とする音声信号を抽出し、インターネットや無線あるいは有線の専用の通信回線を用いて、混合信号から分離された音声信号を別のサーバに出力する。

【００２２】

次に、音声抽出装置の機能について説明する。図２は音声抽出装置の機能ブロック図である。同図に基づいて各機能を説明する。

【００２３】

音声抽出装置１０は、音声音源と非音声音源の信号からなる混合信号をオンラインで分離するために、予め各音声信号の統計状態を音声統計状態推定部６０において推定し、推定結果を音声統計状態記憶部７０に知識として記憶しておく。

【００２４】

非音声区間検出部２０は、信号データのフレーム（信号源の数に対応した信号データをまとめたもので、実際は、ベクトル量として考える）毎に混合信号を入力し、該混合信号について当該フレームの音声または非音声の判定を行う。

【００２５】

非音声統計状態推定部３０は非音声区間検出部２０より音声および非音声判定結果と混合信号を受け取り、非音声区間の場合に混合信号を用いて振幅値の頻度分布を作成する。統計状態は振幅値の頻度を観測数で割ることによって求めることができる。

【００２６】

分離手続き初期化部４０は非音声区間検出部２０より音声または非音声の判定結果と混合信号を、音声抽出部５０より分離手続きの処理方法を受け取る。分離手続き初期化部４０は非音声区間の混合信号を用いて、分離手続きの処理方法の初期手続きを決定する。音声抽出部５０から入力した分離手続きの処理方法と分離手続き初期化部４０によって決定された分離手続きの処理方法との間に、ある数量的に表された一定以上の違いがあれば、音声抽出部５０の分離手続きの処理方法の初期化を行う。

【００２７】

音声抽出部５０は、混合信号、非音声統計状態推定部３０から非音声信号の統計状態、および分離手続き初期化部４０から初期化された分離手続きの処理方法との３つを受け取り、音声統計状態記憶部７０に、知識として予め保持された音声信号源の統計状態と、非音声統計状態推定部３０が算出した非音声音源の統計状態を用いて統計的独立性の尺度を算出し、独立性が高くなる規範で分離手続きの処理方法を更新する。該更新された分離手続きの処理方法を用いて音声信号と非音声信号に分離する。

【００２８】

以上の音声抽出装置の処理手続きの音声通信方式への応用例として、図３を説明する。同図に示すように、音声信号と非音声信号からなる混合信号は、音声抽出装置１の入力手段１と入力手段２によって、２つの混合信号として入力される。これら２つの混合信号は、音声抽出部５０、１００で分離され、音声信号および非音声信号となる。ここまでの処理は音声抽出装置の処理であるが、更に音声抽出装置に音声符号化部１１０、送信部１２０を付加することによって、本発明による音声抽出方法を音声通信方式へ応用可能である。

【００２９】

以下、音声抽出装置の音声抽出部の処理説明に必要となる独立成分分析の具体的な説明を、数式を用いて説明する。独立な音源信号ｓ（ｓは音源信号ベクトル）がある時間定数である混合行列Ａで混合され、混合信号ｘ（混合信号ベクトル）が観測されるものとする。数式で表現すると、$x = As$となる。この式を、以下、観測式と呼ぶことにする。

音源信号ベクトル$s = (s_1(t), \cdots, s_n(t))^T$、

混合信号ベクトル$x = (x_1(t), \cdots, x_n(t))^T$、

混合行列Ａ（ｎ×ｎ行列）、

分離結果の信号ベクトル$y = (y_1(t), \cdots, y_n(t))^T$

とする。

【００３０】

このとき、混合信号ｘを用いて信号の統計的な独立性を手掛りに分離行列Ｗを求め、音源を分離する。数式で表現すると、ｙ＝Ｗｘとなる。この式を、以下、分離式と呼ぶことにする。分離式でｘを分離した（実際には、ベクトルｘの左から行列Ｗを乗算した）際に、ｙが独立となるように分離行列Ｗを求める問題となる。分離行列Ｗは音源間の統計的な独立性が大きくなる方向になるように下記の式で適応的に更新する。

【数１】

$$\Delta W = \eta[I - E[\phi(y)y^T]]W \qquad \cdots (1)$$

ここで、

【数２】

$$\phi(y) = -\left[\frac{\partial}{\partial y1}\log P1(y1),\ldots,\frac{\partial}{\partial yn}\log Pn(yn)\right]^T \qquad \cdots (2)$$

であり、ηは分離更新の大きさを調整する十分小さな正の係数である。

【００３１】

（数１）は、分離結果の信号ベクトルの確率密度関数と分離結果の信号ベクトルの各周辺分布の積との間の相対エントロピー（カルバック・ライブラー情報量など）が計算され、分離行列Ｗの関数である該相対エントロピーが最小化される方向を勾配流の概念を用いて導出される。ここで、重要な事実は、分離結果の確率密度関数を算出する必要はなく、混合信号ベクトルの確率密度関数を導出すればよい。従って、分離結果のベクトルの確率密度関数をいかに混合信号ベクトルから推定すればよいかが問題となる。

【００３２】

以下、音声抽出装置の処理手順に関する３つの実施の形態について説明する。以下では、混合信号の数を２（ｎ＝２）、分離行列Ｗ（２×２行列）の場合について具体的に説明する。

【００３３】

（実施の形態１）

図４は、音声抽出装置の処理の処理手順のフローチャート（その１）である。以下、各工程の説明を同図に基づいておこなう。各音源の確率密度関数を予め知識として記憶する（ステップＳ００１）。ただし、環境雑音等、未知の音源の確率密度関数は当然ながら保持する必要はない。具体的には、例えば、携帯電話を所有する所有人物が、携帯電話を電話として使用する前に所有者の音声信号を携帯電話に記憶させ、その音声の振幅をもとに音声統計状態推定部６０は確率密度関数を作成し、これを携帯電話の音声統計状態記憶部７０に記憶しておく。

【００３４】

次に、分離行列Ｗをランダム値で初期化する（ステップＳ００２）。次に、１フレーム分のＮ個の混合信号を入力する（ステップＳ００３）。例えば、ｘ１＝［ｘ１（０），．．．，ｘ１（ｋ－１）］，ｘ２＝［ｘ２（０），．．．，ｘ２（ｋ－１）］，ここで、ｋは１フレームのサンプル数をあらわす。次に、式ｙ＝Ｗｘにより、分離行列Ｗを用いて音源としての混合信号ベクトルｘを分離する（ステップＳ００４）。式ｙ＝Ｗｘを成分で書くと

【数３】

$$\begin{pmatrix} y1(t) \\ y2(t) \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{pmatrix} x1(t) \\ x2(t) \end{pmatrix} \qquad \cdots (3)$$

となる。

【００３５】
次に、分離信号のうち非音声信号源を特定する（ステップＳ００５）。実際、式ｙ＝Ｗｘで求めた分離信号ｙの成分のうち、非音声に対応する成分を選ぶ。ｙの成分のピッチ相関を算出し、最もピッチ相関が小さいものを非音声に対応する成分とする。非音声に対応する成分がｙ１成分となるように、ｙ、Ｗを並べ替える。たとえば、ｙ２成分が非音声に対応する成分の場合には、ｙ１、ｙ２を入れ替える。これに伴いＷの成分も第一行と第二行を入れ替える。次に、混合信号１の音声または非音声判定を行う。例えば、ピッチ相関を算出し、ピッチ相関＜閾値の場合、非音声と判定し、ピッチ相関≧閾値の場合、音声と判定する（ステップＳ００６）。

【００３６】
次に、混合信号１が非音声と判定された場合は、ステップＳ００８に処理を進める。混合信号１が音声と判定された場合は、ステップＳ０１３に処理を進める（ステップＳ００７）。次に、混合信号２の音声および非音声判定を行う（ステップＳ００８）。例えば、ピッチ相関を算出し、ピッチ相関＜閾値の場合、非音声と判定し、ピッチ相関≧閾値の場合、音声と判定する。次に、混合信号２が非音声と判定された場合は、ステップＳ０１０に処理を進め、混合信号２が音声と判定された場合は、ステップＳ０１３に処理を進める（ステップＳ００９）。

【００３７】
次に、混合信号１、２をパワーで正規化する（ステップＳ０１０）。実際、ｘ１（ｔ）＝ｘ１’（ｔ）／ｐ１およびｘ２（ｔ）＝ｘ２’（ｔ）／ｐ２、（ｔ＝０，．．．，ｋ－１）。ｐ１、ｐ２をそれぞれｙ１、ｙ２のフレームパワーとする。非音声信号源の確率密度関数を算出する（ステップＳ０１１）。ｘ１’（ｔ）およびｘ２’（ｔ）を用いて、正規化した非音声区間の信号の振幅の頻度分布を作成する。ｘ１’（ｔ）およびｘ２’（ｔ）は、－１から＋１の範囲の値となるので、－１から＋１の範囲を適当な数Ｒに分割し、ｒ（ｉ）＝ｑ（ｉ）／$N_{a,1}$として非音声信号源の確率密度関数を算出する。ここで、ｉは０，．．．，Ｒ－１の範囲を動き、ｒ（ｉ）は非音声信号源の確率密度関数、ｑ（ｉ）はｉ番目の区間に入る振幅値の頻度（起動時からの総数）、$N_{a,1}$は振幅値の数（起動時からの総数）である。

【００３８】
次に、非音声信号源の確率密度関数をステップＳ０１１で算出したもので置き換える（ステップＳ０１２）。次に、分離行列Ｗを（数１）によって更新する（ステップＳ０１３）。次に、ステップＳ００４で算出した分離信号を出力する（ステップＳ０１４）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップＳ００３に処理を進める（ステップＳ０１５）。

【００３９】
上述してきたように、本実施の形態１では、非音声区間検出部２０で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができ、これを用いて非音声信号源の確率密度関数を算出できる。これにより非音声信号源が未知である場合や時間変化する場合にも確率密度関数を正確に求めることができ、したがって分離性能の向上が可能となる。

【００４０】
（実施の形態２）
図５は、音声抽出装置の処理の処理手順のフローチャート（その２）である。以下、各工程の説明を同図に基づいておこなう。ステップＳ１０１～ステップＳ１０９までは、実施の形態１と全く同様であるから、説明を省略する。ステップＳ１１０以降を説明する前に、本実施の形態の本質部分である分離行列Ｗの初期値算出の具体的方法について説明をした後、各ステップの説明を行う。

【００４１】
分離行列Ｗの初期値の算出は下記の通りに行う。時刻ｔで音源ｓ１（ｔ）とｓ２（ｔ）が混合行列Ａで混合され、その混合結果ｘ１（ｔ）とｘ２（ｔ）が観測されるものとする。

音源ｓ１は音声信号源、音源ｓ２は非音声信号源であるとする。また混合行列は時間で変化しないと仮定する。数式で表現すれば下記の（数４）となる。

【数４】

$$\begin{pmatrix} x1(t) \\ x2(t) \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} s1(t) \\ s2(t) \end{pmatrix} \quad \cdots (4)$$

ｘ１（ｔ）とｘ２（ｔ）を観測し、元の音源に分離するような分離行列Ｗを推定することで、音源を分離する。数式で表現すれば（数３）となる。

【００４２】

ここで、音声検出手段により、非音声区間を判定すると、その区間では観測される信号には非音声しか含まれていないことと、Ａ＝Ｗ$^{-1}$より、$A_{12}/A_{22} = -W_{12}/W_{11} = \alpha$となる。したがって、Ｗの初期値をＸとすると、Ｘの第一行の成分が、$X_{12}/X_{11} = \alpha$を満たすようにＷの初期値Ｘを算出する。ただし、Ｘの第二行目の成分は自由変数である。

【００４３】

さて、ここで処理手順のステップの説明に戻る。上記説明に基づき、分離行列の初期値Ｘを算出する（ステップＳ１１０）。次に、現在の分離行列Ｗと初期値Ｘについて｜－（W_{12}/W_{11}）－（X_{12}/X_{11}）｜＞ＴＨを評価し、成り立つ場合は、ステップＳ１１２に処理を進め、成り立たない場合は、ステップＳ１１３に処理を進める（ステップＳ１１１）。次に、分離行列を次式で更新する（ステップＳ１１２）。$W_{11} = W_{11}$，$W_{12} = (X_{12}/X_{11}) \times W_{11}$，$W_{21} = W_{21}$，$W_{22} = W_{22}$。次に、ステップＳ４で算出した分離信号を出力する（ステップＳ１１３）。分離行列Ｗを（数１）によって更新する（ステップＳ１１４）。次に、ステップＳ１０４で算出した分離信号を出力する（ステップＳ１１５）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップＳ１０３に処理を進める（ステップＳ１１６）。

【００４４】

上述してきたように、本実施の形態２では、非音声区間検出部２０で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができる。これを用いて分離行列の初期値を算出できる。このことでランダムな初期値よりも正解に近い初期値から計算を出発できること、また環境騒音の変化により分離行列が正解から外れた場合に、正解に近い初期値に設定できるので、正しい分離行列を求め易くなる。

【００４５】

（実施の形態３）

図６は、音声抽出装置の処理の処理手順のフローチャート（その３）である。以下、各工程の説明を同図に基づいておこなう。実施の形態３は、実施の形態１および実施の形態２を組み合わせたものであり、混合信号をパワーで正規化する処理と分離行列の初期値を算出する処理を音声抽出処理に組み込んだものである。

【００４６】

以下、各工程の説明を同図に基づいておこなう。各音源の確率密度関数を予め知識として記憶する（ステップＳ２０１）。ただし、環境雑音等、未知の音源の確率密度関数は当然ながら保持する必要はない。具体的には、例えば、携帯電話を所有する所有人物が、携帯電話を電話として使用する前に所有者の音声信号を携帯電話に記憶させ、その音声の振幅をもとに確率密度関数を作成し、これを携帯電話の記憶部に記憶しておく。次に、分離行列Ｗをランダム値で初期化する（ステップＳ２０２）。

【００４７】

次に、１フレーム分のＮ個の混合信号を入力する（ステップＳ２０３）。例えば、ｘ１＝［ｘ１（０），．．．，ｘ１（ｋ－１）］，ｘ２＝［ｘ２（０），．．．，ｘ２（ｋ－１）］，ここで、ｋは１フレームのサンプル数をあらわす。次に、式ｙ＝Ｗｘにより、分離行列Ｗを用いて音源としての混合信号ベクトルｘを分離する（ステップＳ２０４）。式ｙ

＝Ｗｘを成分で書くと
【数３】

$$\begin{pmatrix} y1(t) \\ y2(t) \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} x1(t) \\ x2(t) \end{pmatrix} \quad \cdots (3)$$

となる。
【００４８】
次に、分離信号のうち非音声信号源を特定する（ステップＳ２０５）。実際、式ｙ＝Ｗｘで求めた分離信号ｙの成分のうち、非音声に対応する成分を選ぶ。ｙの成分のピッチ相関を算出し、最もピッチ相関が小さいものを非音声に対応する成分とする。非音声に対応する成分がｙ１成分となるように、ｙ、Ｗを並べ替える。たとえば、ｙ２成分が非音声に対応する成分の場合には、ｙ１、ｙ２を入れ替える。これに伴いＷの成分も第一行と第二行を入れ替える。次に、混合信号１の音声または非音声判定を行う。例えば、ピッチ相関を算出し、ピッチ相関＜閾値の場合、非音声と判定し、ピッチ相関≧閾値の場合、音声と判定する（ステップＳ２０６）。
【００４９】
次に、混合信号１が非音声と判定された場合は、ステップＳ２０８に処理を進める。混合信号１が音声と判定された場合は、ステップＳ２１６に処理を進める（ステップＳ２０７）。次に、混合信号２の音声および非音声判定を行う（ステップＳ２０８）。例えば、ピッチ相関を算出し、ピッチ相関＜閾値の場合、非音声と判定し、ピッチ相関≧閾値の場合、音声と判定する。次に、混合信号２が非音声と判定された場合は、ステップＳ２１０に処理を進め、混合信号２が音声と判定された場合は、ステップＳ２１６に処理を進める（ステップＳ２０９）。
【００５０】
次に、混合信号１、２をパワーで正規化する（ステップＳ２１０）。実際、ｘ１（ｔ）＝ｘ１’（ｔ）／ｐ１およびｘ２（ｔ）＝ｘ２’（ｔ）／ｐ２、（ｔ＝０，．．．，ｋ－１）。ｐ１、ｐ２をそれぞれｙ１、ｙ２のフレームパワーとする。非音声信号源の確率密度関数を算出する（ステップＳ２１１）。ｘ１’（ｔ）およびｘ２’（ｔ）を用いて、正規化した非音声区間の信号の振幅の頻度分布を作成する。ｘ１’（ｔ）およびｘ２’（ｔ）は、－１から＋１の範囲の値となるので、－１から＋１の範囲を適当な数Ｒに分割し、ｒ（ｉ）＝ｑ（ｉ）／N_{a11}として非音声信号源の確率密度関数を算出する。ここで、ｉは０，．．．，Ｒ－１の範囲を動き、ｒ（ｉ）は非音声信号源の確率密度関数、ｑ（ｉ）はｉ番目の区間に入る振幅値の頻度（起動時からの総数）、N_{a11}は振幅値の数（起動時からの総数）である。
【００５１】
次に、非音声信号源の確率密度関数をステップＳ２１１で算出したもので置き換える（ステップＳ２１２）。次に、分離行列の初期値Ｘを算出する（ステップＳ２１３）。次に、現在の分離行列Ｗと初期値Ｘについて$|-(W_{12}/W_{11})-(X_{12}/X_{11})|>TH$を評価し、成り立つ場合は、ステップＳ２１５に処理を進め、成り立たない場合は、ステップＳ２１６に処理を進める（ステップＳ２１４）。次に、分離行列を次式で更新する（ステップＳ２１５）。$W_{11}=W_{11}$，$W_{12}=(X_{12}/X_{11})\times W_{11}$，$W_{21}=W_{21}$，$W_{22}=W_{22}$。
【００５２】
分離行列Ｗを（数１）によって更新する（ステップＳ２１６）。次に、ステップＳ２０４で算出した分離信号を出力する（ステップＳ２１７）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップＳ２０３に処理を進める（ステップＳ２１８）。
【００５３】
上述してきたように、本実施の形態３では、非音声区間検出部２０で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができるので、これを用いて非音声統

計状態推定部３０で非音声信号源の確率密度関数と分離手続き初期化部４０で分離行列の初期値を算出できる。このことで非音声信号源が未知である場合や時間変化する場合にも確率密度関数を正確に求めることができる。さらに、ランダムな初期値よりも正解に近い初期値から計算を出発でき、また環境騒音の変化により分離行列が正解から外れた場合に、正解に近い初期値に設定できるので、正しい分離行列を求め易くなる。これらにより、分離性能の向上が可能となる。

【００５４】

ところで、本実施の形態では、音声抽出装置の処理手順に関する３つの実施の形態について、混合信号の数を２（ｎ＝２）、分離行列Ｗ（２×２行列）の場合について具体的に説明したが、本発明はこれに限定されるものではなく、ｎが一般の場合においても本発明を適用することもできる。

【００５５】

（付記１）特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、
複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段と、
前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段と、
非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段と、
を備えたことを特徴とする音声抽出装置。

【００５６】

（付記２）前記非音声区間検出手段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする付記１に記載の音声抽出装置。

【００５７】

（付記３）前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする付記１に記載の音声抽出装置。

【００５８】

（付記４）前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする付記１に記載の音声抽出装置。

【００５９】

（付記５）前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段を備えたことを特徴とする付記１に記載の音声抽出装置。

【００６０】

（付記６）特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出プログラムであって、
複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間声検出手順と、
前記非音声区間検出手順により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手順と、
非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手順と、
を備えたことを特徴とする音声抽出プログラム。

【００６１】

（付記７）前記非音声区間検出手順は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする付記６に記載の音声抽出プログラム。

【００６２】

（付記８）前記非音声区間検出手順は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする付記６に記載の音声抽出プログラム。

【００６３】

（付記９）前記音声抽出手順は、前記非音声統計状態推定手順により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする付記６に記載の音声抽出プログラム。

【００６４】

（付記１０）前記非音声統計状態推定手順において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手順を備えたことを特徴とする付記６に記載の音声抽出プログラム。

【００６５】

（付記１１）特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出方法であって、
複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間声検出工程と、
前記非音声区間検出工程により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定工程と、
非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出工程と、
を備えたことを特徴とする音声抽出方法。

【００６６】

【発明の効果】

以上説明したように、請求項１の発明によれば、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、検出された非音声区間の情報を用いて非音声の統計状態を推定し、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の声を抽出することとしたので、時間変化する環境雑音を伴う混合信号から目的とする音声信号を抽出することが可能な音声抽出装置が得られるという効果を奏する。

【００６７】

また、請求項２の発明によれば、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することとしたので、単位時間毎に変化する環境雑音を非音声信号として抽出でき、該抽出結果を用いて、非音声信号の統計状態を推定することが可能な音声抽出装置が得られるという効果を奏する。

【００６８】

また、請求項３の発明によれば、単位時間毎に混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうこととしたので、単位時間毎に混合信号の中から環境雑音を抽出し、単位時間毎に環境雑音の統計状態を推定することが可能な音声抽出装置が得られるという効果を奏する。

【００６９】

また、請求項４の発明によれば、推定された時間によって変動する非音声の統計状態と知識としての音声に関する統計状態との統計的な独立性を用いて、非音声の統計状態と音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、単位時間毎に逐次更新

する分離手続きを行うこととしたので、混合信号の統計状態から音声信号の統計状態を分離でき、よって混合信号から目的とする音声信号を抽出することが可能な音声抽出装置が得られるという効果を奏する。

【００７０】

また、請求項５の発明によれば、混合信号の分離手続きを初期化することとしたので、手続きの初期状態から安定であり確度の高い音声信号の抽出が可能な音声抽出装置が得られるという効果を奏する。

【図面の簡単な説明】

【図１】本発明の概略構成を示す図である。

【図２】音声抽出装置の機能ブロック図である。

【図３】本発明の音声通信方式への適用例の図である。

【図４】音声抽出装置の処理手順のフローチャート（その１）である。

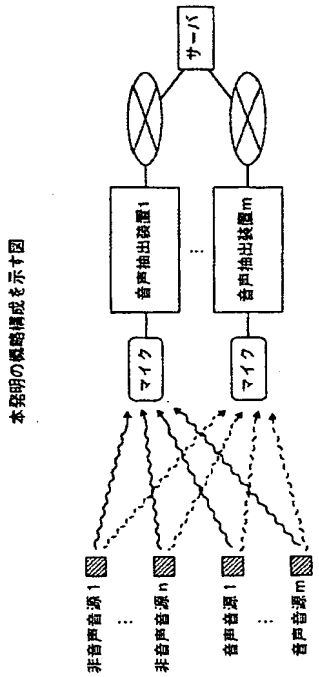
【図５】音声抽出装置の処理手順のフローチャート（その２）である。

【図６】音声抽出装置の処理手順のフローチャート（その３）である。

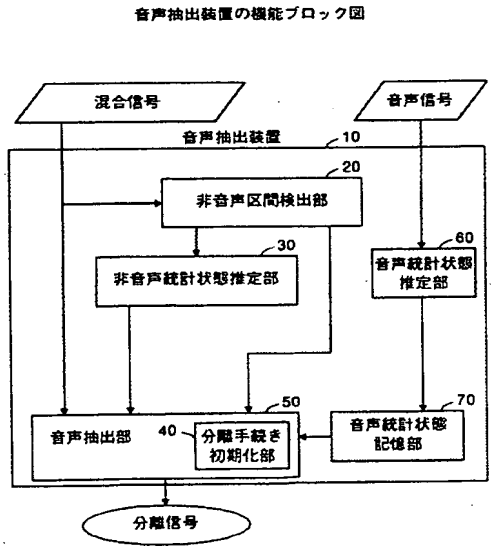
【符号の説明】

１０　音声抽出装置
２０　非音声区間検出部
３０　非音声統計状態推定部
４０　分離手続き初期化部
５０，１００　音声抽出部
６０　音声統計状態推定部
７０　音声統計状態記憶部
１１０　音声符号化部
１２０　送信部

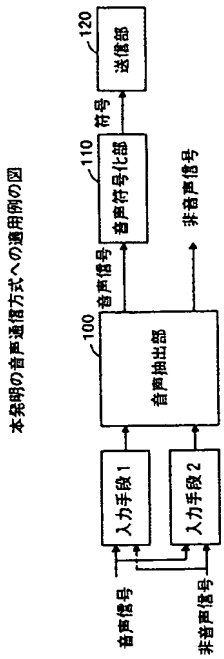
【図１】



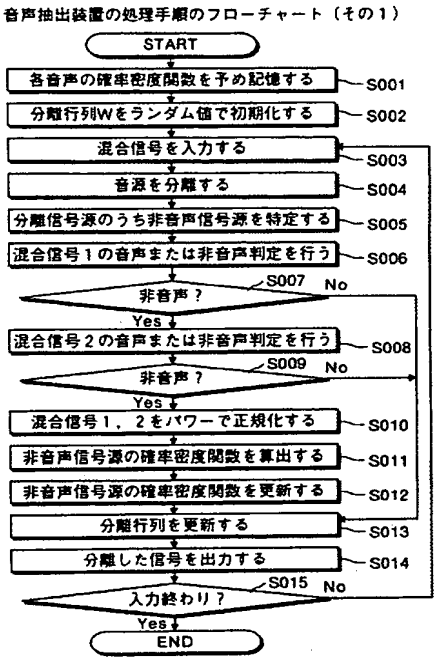
【図２】



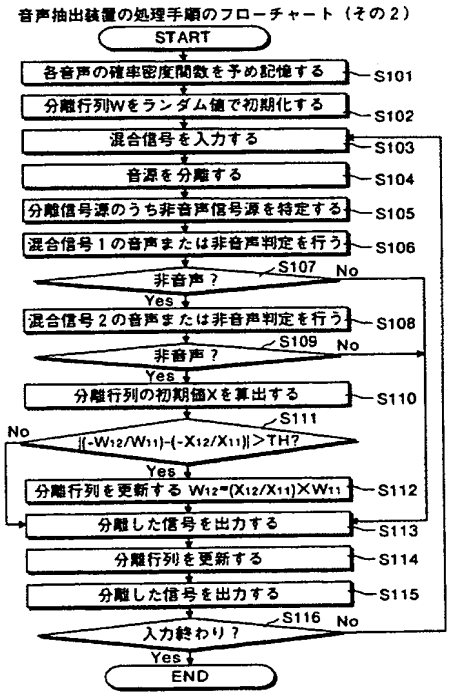
【図３】


本発明の音声通信方式への適用例の図
120
送信部
符号
110
音声符号化部
非音声信号
音声情報
100
音声抽出部
入力手段１
入力手段２
音声信号
非音声信号

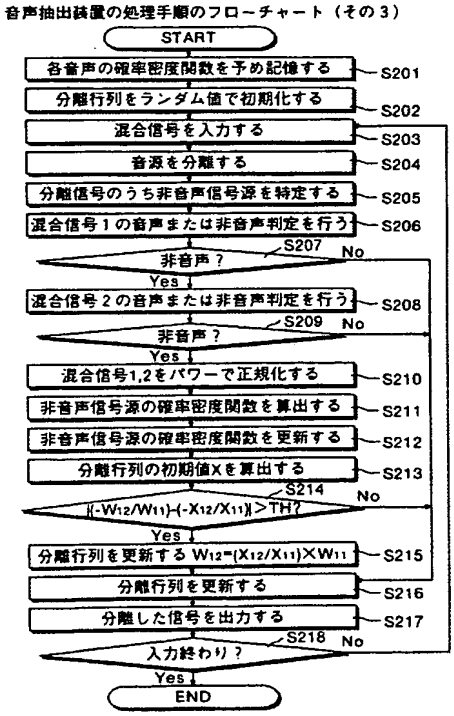
【図４】


音声抽出装置の処理手順のフローチャート（その１）
START
各音声の確率密度関数を予め記憶する
S001
分離行列Wをランダム値で初期化する
S002
混合信号を入力する
S003
音源を分離する
S004
分離信号源のうち非音声信号源を特定する
S005
混合信号１の音声または非音声判定を行う
S006
非音声？
S007
No
Yes
混合信号２の音声または非音声判定を行う
S008
非音声？
S009
No
Yes
混合信号１，２をパワーで正規化する
S010
非音声信号源の確率密度関数を算出する
S011
非音声信号源の確率密度関数を更新する
S012
分離行列を更新する
S013
分離した信号を出力する
S014
入力終わり？
S015
No
Yes
END

【図５】


音声抽出装置の処理手順のフローチャート（その２）
START
各音声の確率密度関数を予め記憶する
S101
分離行列Wをランダム値で初期化する
S102
混合信号を入力する
S103
音源を分離する
S104
分離信号源のうち非音声信号源を特定する
S105
混合信号１の音声または非音声判定を行う
S106
非音声？
S107
No
Yes
混合信号２の音声または非音声判定を行う
S108
非音声？
S109
No
Yes
分離行列の初期値Xを算出する
S110
S111
No
|(-W12/W11)-(-X12/X11)|>TH?
Yes
分離行列を更新する W12=(X12/X11)×W11
S112
分離した信号を出力する
S113
分離行列を更新する
S114
分離した信号を出力する
S115
入力終わり？
S116
No
Yes
END

【図６】


音声抽出装置の処理手順のフローチャート（その３）
START
各音声の確率密度関数を予め記憶する
S201
分離行列をランダム値で初期化する
S202
混合信号を入力する
S203
音源を分離する
S204
分離信号のうち非音声信号源を特定する
S205
混合信号１の音声または非音声判定を行う
S206
非音声？
S207
No
Yes
混合信号２の音声または非音声判定を行う
S208
非音声？
S209
No
Yes
混合信号1,2をパワーで正規化する
S210
非音声信号源の確率密度関数を算出する
S211
非音声信号源の確率密度関数を更新する
S212
分離行列の初期値Xを算出する
S213
S214
No
|(-W12/W11)-(-X12/X11)|>TH?
Yes
分離行列を更新する W12=(X12/X11)×W11
S215
分離行列を更新する
S216
分離した信号を出力する
S217
入力終わり？
S218
No
Yes
END

フロントページの続き

(51)Int.Cl.'	F I	テーマコード（参考）
G10L 21/02		

(72)発明者 松原 光良
福岡県福岡市博多区博多駅前三丁目２２番８号 富士通九州ディジタル・テクノロジ株式会社内

Fターム(参考) 5D015 AA03 DD03 EE04